Read Spectrometer Dataset Example

This example shows how to use function `pnnl_read_spectrometer_dataset` to load spectrometer datasets from multiple files.

Supporting Information

A Practical Guide to Chemometric Analysis of Optical Spectroscopic Data

Hope E. Lackey,[1,2] Rachel L. Sell,[1] Gilbert L. Nelson,[3*] Thomas A. Bryan,[4*] Amanda M. Lines,[1,2*] Samuel A. Bryan,[1,2*]

[1] Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA 99352

[2] Washington State University, Department of Chemistry, Pullman WA 99164

[3]College of Idaho, Department of Chemistry, 2112 Cleveland Blvd, Caldwell, ID 83605

[4] The MathWorks, 3 Apple Hill Drive, Natick, MA 01760-2098

*Email: sam.bryan@pnnl.gov Phone 1 509 375 5648; orcid.org/0000-0002-8826-0880

*Email: tbryan@mathworks.com Phone 1 508 647 7669

*Email: amanda.lines@pnnl.gov Phone: 1 509 375 5689

*Email: gnelson@collegeofidaho.edu Phone: 1 208 459 5241

Data Organization

The spectrometer data is organized in the following way.

- All files from a single dataset are in one directory.
- The data in each file are row oriented.
- Each file contains one spectrum.
- The first row in each file is the "x" data (Wavenumbers).
- All wavenumbers are the same for all of the files in a directory.
- The second row in each file is the "y" data (Absorbance spectrum).

Function `pnnl_read_spectrometer_dataset` has the following syntax that you can see by using the help command on the function (as you can with any MATLAB function).

```
help pnnl_read_spectrometer_dataset
```

```
  pnnl_read_spectrometer_dataset Read spectrometer dataset

    [AbsorbanceMatrix, Wavenumbers] = ...
        pnnl_read_spectrometer_dataset(directory_name, file_extension)

    reads all files in a directory specified by directory_name with
    file extension specified by file_extension.
```

```
If no directory name is specified, or if the directory name is an
empty string, then the files are read from the current directory.

If no file extension is specified, or if file extension is an empty
string, then a file extension of 'txt' is used.

[AbsorbanceMatrix, Wavenumbers] = pnnl_read_spectrometer_dataset
with no additional input arguments reads all files in the current
directory with file extension 'txt'.

Files whose names start with '._' are ignored.  They are assumed to
be file attributes created by Mac computers.

Example:

  training_data_directory_name = ...
      fullfile('pnnl_spectrometer_data', ...
               'pnnl_napalm_training_dataset');
  file_extension = 'txt';
  [A_train, Wavenumbers] = ...
      pnnl_read_spectrometer_dataset(training_data_directory_name, ...
                                     file_extension);
  plot(Wavenumbers,A_train)
  xlabel('Wavenumber (cm^{-1})')
  ylabel('Absorbance')
  title('Training Spectra')
```

The files in this example are the spectrometer dataset files that were used to create `A_train`, `A_unknown`, and `Wavenumbers` that are used in the rest of PNNL Chemometric Toolbox, and are identical to the data you get when you load `pnnl_napalm_data.mat` in the MATLAB Workspace.

If your data is in a different format, then you can copy `pnnl_read_spectrometer_dataset.m` and modify it for your own purposes.

Clear variables from the workspace before beginning the example.

```
clearvars
```

Specify Directories Containing Training and Unknown Dataset files

All training data is in one directory and all unknown data is in another directory. All of the files in each directory are assumed to be in the same dataset.

Change the locations of the directories to where your own data is stored.

Both datasets are contained in a directory named `pnnl_spectrometer_data` that is a sub-directory of the `pnnl_chemometric_toolbox` directory.

The training dataset is contained in a sub-directory of `pnnl_spectrometer_data` named `pnnl_napalm_training_dataset`.

Use `fullfile` to specify a path to a directory with a subdirectory so the path names work on Windows, Linux, and Mac. If you are in the directory that contains the dataset files, then let the directory name be an empty string `''`.

```
training_data_directory =
fullfile('pnnl_spectrometer_data','pnnl_napalm_training_dataset')
```

```
training_data_directory =
'pnnl_spectrometer_data/pnnl_napalm_training_dataset'
```

The training data directory has 20 text files containing one spectrum in each file.

```
dir(training_data_directory)
```

```
.              ..             xy_train_01.txt  xy_train_02.txt  xy_train_03.txt  xy_train_04.txt  xy_t
```

The unknown dataset is contained in a sub-directory of `pnnl_spectrometer_data` named `pnnl_napalm_unknown_dataset`.

```
unknown_data_directory =
fullfile('pnnl_spectrometer_data','pnnl_napalm_unknown_dataset')
```

```
unknown_data_directory =
'pnnl_spectrometer_data/pnnl_napalm_unknown_dataset'
```

The unknown directory has 12 text files containing one spectrum in each file.

```
dir(unknown_data_directory)
```

```
.                ..               xy_unknown_01.txt  xy_unknown_02.txt  xy_unknown_03.txt  xy_unknown_
```

Specify the File Extension

All of the files in the dataset were saved with a "txt" extension by the spectrometer.

```
file_extension = 'txt';
```

Load the Datasets

Read the spectrometer dataset for the training data. The absorbance data is the first output. The wavenumber vector is the second output.

```
[A_train, Wavenumbers] =
pnnl_read_spectrometer_dataset(training_data_directory, file_extension);
```

Read the spectrometer dataset for the unknown data. The absorbance data is the first outout. The wavenumbers are the same for the unknown data and the training data, so you don't need to use the second output in this case.

```
A_unknown = pnnl_read_spectrometer_dataset(unknown_data_directory,
file_extension);
```

Examine the Data

List all the variables in the workspace.

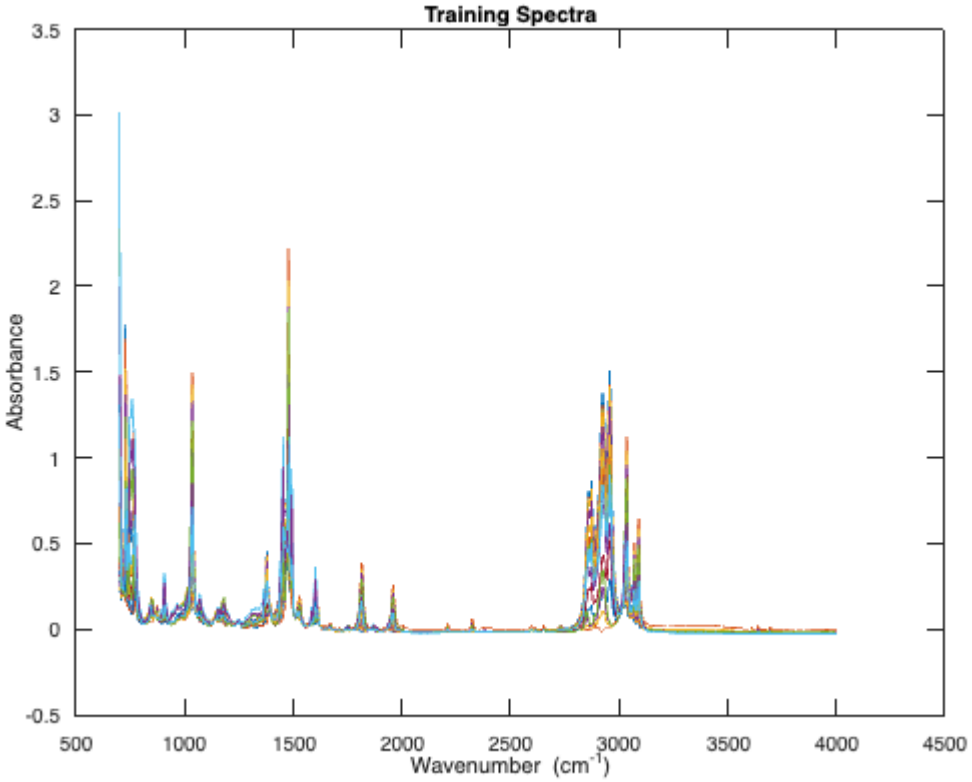
```
whos
```

```
  Name                        Size              Bytes  Class     Attributes

  A_train                    20x1713            274080  double
  A_unknown                  12x1713            164448  double
  Wavenumbers                 1x1713             13704  double
  file_extension              1x3                    6  char
  training_data_directory     1x51                 102  char
  unknown_data_directory      1x50                 100  char
```

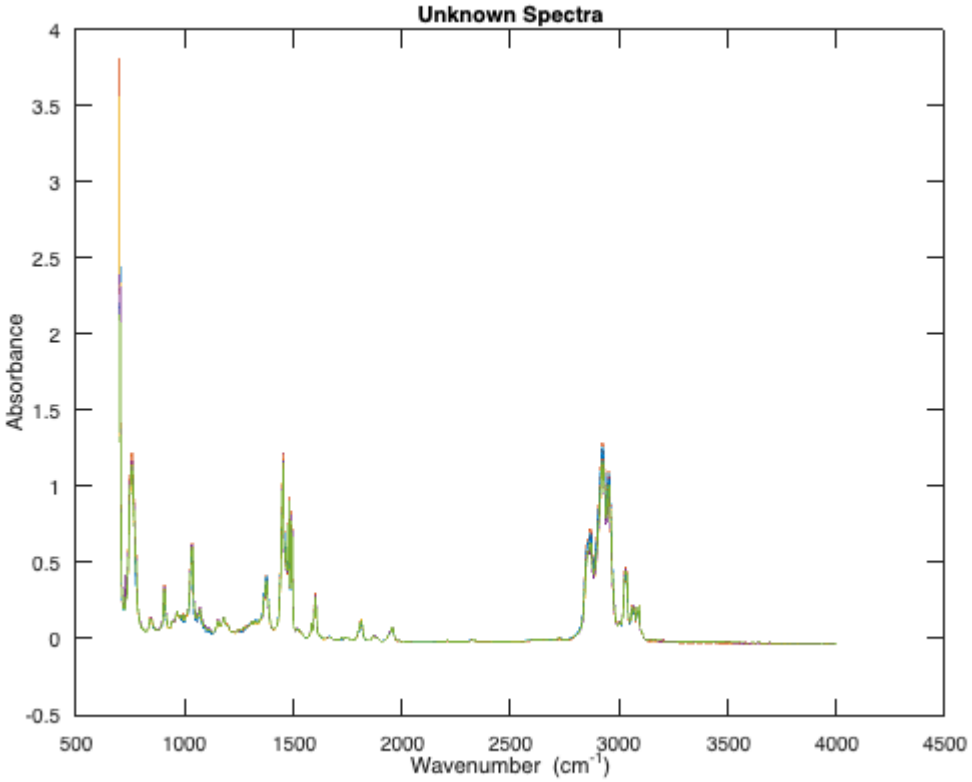
Plot the training spectra.

```
plot(Wavenumbers,A_train)
xlabel('Wavenumber (cm^{-1})')
ylabel('Absorbance')
title('Training Spectra')
```



Plot the unkown spectra.

```
plot(Wavenumbers,A_unknown)
xlabel('Wavenumber (cm^{-1})')
ylabel('Absorbance')
title('Unknown Spectra')
```



Disclaimer